

Comparative Analysis of Three Bovine Genomes September 10, 2019

Sample to Insight -

QIAGEN Aarhus · Silkeborgvej 2 · Prismet · 8000 Aarhus C · Denmark Telephone: +45 70 22 32 44 · www.qiagenbioinformatics.com · ts-bioinformatics@qiagen.com



#### **Comparative Analysis of Three Bovine Genomes**

This tutorial takes you through some of the tools for identifying species-specific variants. As an example we use *Bos taurus* (cow), *Bos indicus* (zebu), and *Bison bison* (bison) data sets consisting of single GA and GAII exome sequencing short reads from an Illumina sequencer. For this tutorial, you will need to use CLC Genomics Workbench Version 9.5 or higher.

The goal is to identify zebu-specific (*Bos indicus*) exome variations that are not present in the bison (*Bison bison*) or bovine genome (*Bos taurus*) [Cosart et al., 2011]. Furthermore, we attempt to link the variations to altered pathways using the gene ontology tool that is built into the *CLC Genomics Workbench*.

#### Downloading the reference data

Download the bovine genome with annotations using the Reference Data Manager (1) found in the upper right corner of the Workbench (figure 1). Under the **Download Genome** tab (2), select the **Bos taurus - Hereford** data (3). Choose to "Download genome sequence" (4) and check the "Genome Annotations" item (5) to get corresponding annotation tracks. Click on **Download** (6). You can check the Download process at the bottom of the wizard.

						Workspace Plugins Re	ferences Download Workflow
Gr Manage Reference Data							×
Download Genomes Robe Reporting Repo		Imported Data Imported Reference Data				ree space in CLC_Ref	erence Data: Locally v erences location: 56.94 GB y folder location: 56.94 GB
Animal (mammals)     Bos taurus - Hereford     Select to download latest data for Bos taurus - Hereford	De	Bos taurus - Hereford		4			•
Canis familiaris     Select to download latest data for Canis familiaris		Download genome seque Use existing genome sec					
Equus caballus Select to download latest data for Equus caballus							
Homo sapiens - hg 18 Select to download latest data for Homo sapiens - hg 18	. 0	Name	5	Version	Provider	Download size	for selection: 778.91
Homo sapiens - hg 19 Select to download latest data for Homo sapiens - hg 19		/ Sequence		UMD3.1	Ensembl	766.0 M	18
Select to download latest data for Homo sapiens - hg19		Genome Annotations		UMD3.1.94	Ensembl	12.9 M	3
Homo sapiens - hg38 Select to download latest data for Homo sapiens - hg38		Variants		Downle	Ensembl ad Genome	1.61 GE	Gance
Mus musculus     Select to download latest data for Mus musculus	Pr	evious downloads		COMIN		6	Download
Rattus norvegicus Select to download latest data for Rattus norvegicus		Date	Size		Versions	Folder	
Sus scrofa Select to download latest data for Sus scrofa							Delete
Animal (others)	* (				III		- F
Help							Close

Figure 1: The Reference Data Manager is found in the upper right corner of the Workbench.

The reference sequence and annotations tracks are now saved in the CLC\_References | Genomes | Bos\_taurus\_Hereford folder accessible from the Navigation Area.

#### **Importing reads**

1. Download the read files using the tool:

#### Download | Search for Reads in SRA ( 🚑 )

2. This will bring up the dialog shown in figure 2. Search for one of the run with the following accession number SRR307349.



	Copy Paste Delete										Workspace Plugins References D	Sownload Workflo
🖉 SR	A Search ×									Æ	Search for Reads in SRA	
All Fi	elds 👻 SRI	R307349							×	S.	Search for Sequences at NCBI	Ctrl+B
h	4								P Start search	8	Search for PDB Structures at NCBI	
				Add sear	ch parameter				By Start search	$\mathcal{Q}$	Search for Sequences in UniProt	Ctrl+Shift+
										1 mm		
#	Run Accession	Study Accession	Scientific Name	Download Size	Paired	Read Orientation	Average Length	Spots	PubMed			

Figure 2: Search for a particular run using the Search for Reads in SRA tool.

3. The search output one run. To find the other reads generated during this study, right click on the row of the search output table and choose the option "search for..." and select SRP007095, i.e., the accession number of the study (figure 3).

	ields 🚽 SRF	307349									E	Search Par	ameters Settings
		(307315										Column width	. [
						Add sear	rch parameter	]		[	🖞 Start search	Show column	[
ŧ	Run Accession	Study Accessi	on	Scientific Na	me	Download Size	Paired	Read Orientation	Average Length	Spots	PubMed	SRA Preview	
	SRR307349	SRP007095	File Edit View Toolb Show Searc	► h for ►	Bos tau SRP0070 SRS2118 SRX0730	rus 095 SRP007095	No	Uninown	3e	4,127,4	77, <u>21729323</u>	a Bos taurus Genomic DN using probe	ons were resequenced in 3 individuals: s, a Bos indicus, and a Bison bison. JA was entriched for an exome subset s on a microarray, probing for 16,131 nces across the X and the 29 somatic

Figure 3: Search for more runs from the same experiment.

4. You have now 6 different runs in the search output table. To minimize the tutorial download and analysis time, select the 3 runs that have the lowest Download Size and click on Download Reads and Metadata (figure 4).

ALE	ields 🗸								×	Search Parameter	ers Settings
	,									Column width	-
				Add search parar	neter			් පී s	tart search		Automatic 👻
#	Run Accession	Study Accession	Scientific Name	Download Size (MByt	Paired Re	ead Orientation	Average Len	Spots	PubMed	Show column	
	SRR 307349	SRP007095	Bos taurus		No Uni	known		4,127,477	21729323	SRA Preview	
	SRR.307350	SRP007095							21729323	Title:	
	SRR.307351	SRP007095				iknown		7,163,414	21729323		
4	SRR.308478	SRP007095	Bison bison	526	No Uni	known	36	24,687,120	21729323	GAI, Bos taurus	
5	SRR 308479	SRP007095	Bos indicus	520	No Uni	known	36	23,614,360	21729323	Abstract:	uara racaquanaad in 2 individuala:
6	<u>SRR 308 480</u>	SRP007095	Bos taurus	571	No Un	iknown	36	24,239,424	21729323	a Bos taurus, a B Genomic DNA wa using probes on exon sequences	vere resequenced in 3 individuals: los indicus, and a Bison bison. as entriched for an exome subset a microarray, probing for 16,131 across the X and the 29 somatic
		bownload Reads ar	nd Metadata	Show Metadata for Selec	tion Total nu	umber of experiment	s: 6 more	<b>.</b>		chromosomes.	
E 🖸	2 🗳									- 그 리	? ⊫

Figure 4: Download in one click the runs and associated metadata.

- 5. In the download dialog that pops up, choose to discard read names and quality scores.
- 6. Save the data in the folder you created for this tutorial.
- 7. For clarity in subsequent analysis, rename the reads with the corresponding species name (right click on the read file in the Navigation Area).

When all data are downloaded you can see the following files in the **Navigation Area** as shown in figure 5.

We are now ready to start analyzing the data.



Comparative analysis tutorial reads SRA MetadataTable Bos taurus Bos indicus Ference data Ros taurus (Hereford) sequence Bos taurus (Hereford)\_Transcript Bos taurus (Hereford)\_CDS Bos taurus (Hereford)\_CDS Bos taurus (Hereford)\_Exon Bos taurus (Hereford)\_Gene Bos taurus (Hereford)\_Five\_prime\_utr Bos taurus (Hereford)\_Five\_prime\_utr Bos taurus (Hereford)\_Five\_prime\_utr Bos taurus (Hereford)\_Five\_prime\_utr Bos taurus (Hereford)\_Five\_prime\_utr

Figure 5: The downloaded reference genome data.

#### Map the reads and detect variants

The data used in this tutorial do not contain adapters, so we do not need to run a trimming step. Thus, we proceed directly to mapping the reads to the reference.

We will map the reads to the reference genome using the "Batch" mode, which means that the three sets can be run in parallel (simultaneously).

1. To map reads to the reference, go to the toolbox:

#### Toolbox | Resequencing Analysis (🚮) | Map Reads to Reference (式)

In the first wizard step (figure 6), check the **Batch** box, then select the folder that holds the reads. Do not forget to check the batch box as it would generate a single read mapping with the reads of all three species. Click **Next**.

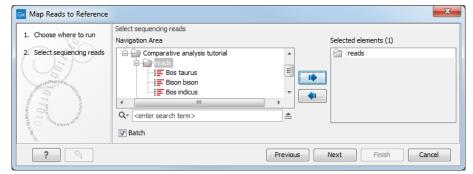


Figure 6: Run the analysis in batch mode to generate one read mapping per species.

- 2. You should now see the batch overview window (figure 7). In this wizard step you can see the reads that were contained in the selected folder. In this case we will use all the reads in the analysis so you can just click **Next**.
- 3. In the next wizard window (figure 8), click on the browse icon () under **References** to select the *Bos taurus* (Hereford) reference sequence. Press OK. Leave the Reference masking parameters as set by default to "No masking" and click **Next**.
- 4. In the next wizard step, you can specify the **Mapping options**. We will use the default settings and click **Next**.
- 5. Select the options "Create reads track", "Create report" and "Save in specified location" using the option to "Create subfolders per batch unit". Specify where you want to save the data, for example a new folder called "Analysis" and click on the button labeled **Finish**.



1. Choose where to run	Batch overview Units	Contents	
2. Select sequencing reads	Bos taurus		
3. Batch overview	Bison bison		
	Bos indicus		
C SO ANTIMAL			
LEP			
8			
Sold Artis And Artis			
1 0			 
Tan Juni	Only use elements containing:		
	Exclude elements containing:		
LEATDANNIE STATE			6 elements in to

Figure 7: Failing to see this dialog means you forgot to check the batch option previously!

Gx Map Reads to Reference			23	
1. Choose where to run	References			
2. Select sequencing reads	References 🎇 Bos taurus (H	lereford) sequence	Ø	
<ol> <li>Batch overview</li> <li>References</li> </ol>	Reference masking	Select single genome track or reference sequences		<u></u>
100 1 10 10 10 10 10 10 10 10 10 10 10 1	Exclude annotated     Include annotated only     Masking track	Navigation Area Comparative analysis tutorial reads Bison bison Bison bison reference data Correct turus (Hereford) sequence (< turus (Hereford) sequence		Selected elements (1)
		OK	Cancel	

Figure 8: Select the Bos taurus (Hereford) reference sequence.

Open the mapping summary reports from the **Navigation Area**. 100% of reads are mapped successfully in all three analyses. We can then proceed to a local realignment of the reads to improve the quality of the read mappings we have just generated.

1. To run the local realignment tool go to the toolbox:

#### Toolbox | Resequencing Analysis () | Local Realignment

- 2. Again we choose to run the analysis in **batch mode**, which allows us to analyze all samples in parallel. Select all the read mapping tracks by selecting the top folder as shown in figure 9 and click **Next**.
- 3. The next dialog in which you can see the content of the selected folder confirms that you are working in batch mode. Click **Next**.
- 4. In the next wizard step, use the default settings (we will run the local realignment without a guidance track). Click on the button labeled **Next**.



2. Select read mapping reads reference data reference data	▲ III Analysis
---	----------------

Figure 9: Run local realignment for all species at the same time by using the batch function.

5. Click on **Create reads track** (see figure 10) and choose to **Save in input folder**. Click on **Finish**.

Gx Local Realignment	Σ.
Choose where to run     Select read mapping     Batch overview     Realignment settings	Result handling         Output options
5. Result handling	Result handling Open Save in input folder Save in specified location Create subfolders per batch unit Log handling Open log
? 🔊	Previous Next Finish Cancel

Figure 10: Use the default settings and choose to save the results.

We are now ready to identify the variants. Note that we are generating here a variant track for all three species, when in fact we will use only the *Bos indicus* variant track in subsequent analyses.

1. Go to the toolbox:

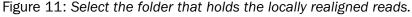
### Toolbox | Resequencing Analysis () | Variant Detection () | Basic Variant Detection ()

- 2. Using the batch function, select the top folder as in figure 11.
- 3. Click through the next wizard steps (by clicking on the button labeled Next) using the default settings. If you have changed some of the settings, you can always go back to the default settings by pressing the **Reset** button in the lower left corner, which will return parameters to default settings.
- 4. In the last wizard step choose to **Create reads track** and **Save in input folder** before clicking on **Finish** (figure 12).

Take a look at the results, which should be organized as shown in figure 13.



Gx Basic Variant Detection		X
1. Choose where to run 2. Select read mappings	Select read mappings Navigation Area Comparative analysis tutorial Freads Freference data Fref	Selected elements (1)
?		Previous Next Finish Cancel



Gx Basic Variant Detection	X
1. Choose where to run	Result handling
2. Select read mappings	Output options
3. Basic Variant Parameters	Create annotated table
4. General filters	
5. Noise filters	Result handling
6. Result handling	⊙ Save
Contraction of the second seco	Log handling
TT TO TO T	Open log
Help Reset	Previous Next Finish Cancel

Figure 12: Create a reads track and save the results in their respective input folders.

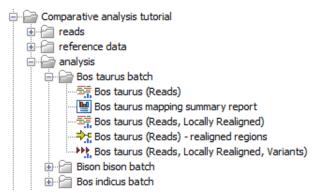


Figure 13: Example of how all outputs have been saved to the batch specific folder created during the first batch analysis.

Let's take a look at the one of the variant tracks. Open the **Bos taurus (Reads) - locally realigned (Variants)** track by double-clicking on the name in the **Navigation Area**. Use the lower left corner buttons to open the variant track in table view. Open the table in split view by holding down the Ctrl key (Cmd on Mac) while clicking on the table icon (see the red arrow in figure 14). You will be able to see the variant track and the variant table in the same window.

The variant detection tool identifies around 8,000 variants in the bovine genome, about 13,000 in zebu, and about 23,000 in bison.



						Show	w more tracks tog	ether:	Create Track List
		20,0	00,000 I	40,000,000 I	60,000,000 I 68,	80,000,0 923,194 I	000 10	0,000,000 I	120,000,000 I
Bos taurus (F Locally Real					- L -				,
lu 🔠 🖸 🕻								_	
Rows: 8,26	i0 Tabl	e view: Bos taur			-				Filter
Rows: 8,26	50 Table Region	Туре	Reference	Allele	Reference	Length	Count	Coverage	Frequency
Rows: 8,26	60 Table Region 438249	Type SNV	Reference G	A	No	1	5	12	Frequency 41.67
Rows: 8,26	i0 Tabl Region 438249 438249	Type SNV SNV	Reference G G	A G	No Yes	1	5 7	12	Frequency 41.67 58.33
Rows: 8,26	60 Table Region 438249 438249 438254	Type SNV SNV SNV	Reference G G T	A G A	No Yes No	1	5 7 8	12 12 16	Frequency 41.67 58.33 50.00
Rows: 8,26	io Table Region 438249 438249 438254 438254	Type SNV SNV SNV SNV SNV	Reference G G T T	A G A T	No Yes No Yes	1 1 1 1	5 7 8 8	12 12 16 16	Frequency 41.67 58.33 50.00 50.00
Rows: 8,26	Region           438249           438249           438254           438254           438254           438254	Type SNV SNV SNV SNV Deletion	Reference G G T T T T	A G A T	No Yes No Yes No	1 1 1 1 1	5 7 8 8 7	12 12 16 16 19	Frequency 41.67 58.32 50.00 50.00 36.84
Rows: 8,26	Region         Table           438249         438249           438254         438254           438254         438257           438257         438257	Type SNV SNV SNV SNV Deletion SNV	Reference G G T T T T T	A G A T - T	No Yes No Yes No Yes	1 1 1 1 1 1	5 7 8 8 7 12	12 12 16 16 19 19	Frequency 41.67 58.33 50.00 50.00 36.84 63.16
Rows: 8,26	Region           438249           438254           438254           438257           438257           438257	Type SNV SNV SNV SNV Deletion SNV SNV	Reference G G T T T T C	A G A T - T T	No Yes No Yes No Yes No	1 1 1 1 1 1 1	5 7 8 8 7 12 8	12 12 16 16 19 19 19	Frequency 41.67 58.33 50.00 50.00 36.84 63.16 66.67
Rows: 8,26	Region           438249           438254           438254           438257           1563318           1577741	Type SNV SNV SNV SNV Deletion SNV SNV SNV	Reference G G T T T T T C G	A G A T - T T T	No Yes No Yes Yes No No	1 1 1 1 1 1 1 1 1	5 7 8 8 7 12 8 10	12 12 16 16 19 19 19 12 10	Frequency 41.67 58.33 50.00 36.84 63.16 66.67 100.00
Rows: 8,26	Region           438249           438254           438254           438257           438257           438257	Type SNV SNV SNV SNV Deletion SNV SNV SNV SNV	Reference G G T T T T C G G C	A G A T - T T T T	No Yes No Yes No Yes No	1 1 1 1 1 1 1 1 1 1 1	5 7 8 8 7 12 8 10 9	12 12 16 16 19 19 19	Frequency 41.67 58.33 50.00 36.84 63.16 66.67 100.00 90.00
Rows: 8,26 Chromosome	Region           438249           438249           438254           438254           438257           1563318           1577741           1577769	Type SNV SNV SNV SNV Deletion SNV SNV SNV	Reference G G T T T T T C G	A G A T - T T T	No Yes No Yes Yes No No No	1 1 1 1 1 1 1 1 1	5 7 8 8 7 12 8 10 9 9	12 12 16 16 19 19 12 10 10	Frequency 41.67 58.33 50.00 50.00 36.84 63.16 66.67 100.00 90.00 100.00
Rows: 8,26	Region           438249           438249           438254           438254           438257           1563318           1577741           157759           1582959	Type SNV SNV SNV SNV Deletion SNV SNV SNV SNV SNV SNV	Reference G G T T T C C G C A	A G A T - T T T G	No Yes No Yes No Yes No No No No	1 1 1 1 1 1 1 1 1 1 1 1	5 7 8 8 7 12 8 10 9 9	12 12 16 16 19 19 19 12 10 10 10	Frequency 41.67 58.33 50.00 36.84 63.16 66.67 100.00 90.00 100.00 38.46
	Region           438249           438249           438254           438257           1563318           1577741           1577769           158259           1584650	Type SNV SNV SNV SNV SNV Deletion SNV SNV SNV SNV SNV SNV SNV	Reference G G T T T C G G C C A C	A G A T T T T G T	No Yes No Yes No No No No No No	1 1 1 1 1 1 1 1 1 1 1 1	5 7 8 8 7 12 8 10 9 9 17 7 5	12 12 16 16 19 19 12 10 10 10 17 7 13	Frequency 41.67 58.33 50.00 36.84 63.16 66.67 100.00 90.00 100.00 38.46

Figure 14: Open the variant table in split view.

#### **Identify Species-specific Variants**

We are now going to identify variants that are unique for the zebu species with the tool **Identify known mutations from sample mapping**.

1. Go to:

Toolbox | Resequencing Analysis () | Identify known mutations from sample. mapping ()

- 2. In the first wizard dialog, select the read mappings of *Bos taurus* and *Bison bison*. Click **Next**.
- 3. Select the zebu variant track **Bos indicus (Reads) locally realigned (Variants)**. Leave the other parameters at their default value and click on the button labeled **Next**.
- 4. Create individual tracks and an overview track. Choose to Save them in a specified location (for example a new folder called zebu specific variants), and click on **Finish**.
- 5. Now open the file **Bos indicus (Reads, Locally Realigned, Variants) (MutationTest overview)**. We want to filter away from the table the variants that were also detected in the two other species, and keep only the ones that were not (the detection value is set to no). Open the filter and select "Bos taurus detection" in the first field, "=" in the second, and type "no" in the third field. Add an additional filter that you set up as you did the previous one but for "Bison bison detection" (see figure 15).

			r	us (Reads, Locall on (Reads, Locall			•	no		× Filter
	Frequency	Bos ta	aurus (	Bos taurus (	Bos taurus (	Bos taurus (Re	Bison bison	Bison bison	Bison bison	Bison bison (Re
15	60.00	no		no	0.00	Unknown	yes	no	6.25	Unknown
23	56.52	yes		no	0.00	Unknown	yes	no	0.00	Unknown
16	50.00	no		no	0.00	Unknown	yes	no	0.00	Unknown
24	100.00	yes		no	11.11	Unknown	yes	no	0.00	Unknown
15	100.00	yes		no	0.00	Unknown	yes	no	0.00	Unknown
10	90.00	no		no	0.00	Unknown	no	no	0.00	Unknown
16	56.25	yes		no	0.00	Unknown	yes	no	0.00	Unknown
12	66.67	no		no	0.00	Unknown	no	no	0.00	Unknown
10	50.00	no		no	0.00	Unknown	no	no	0.00	Unknown
24	87.50	yes		no	0.00	Unknown	yes	no	0.00	Unknown
24	41.67	yes		no	0.00	Unknown	no	no	0.00	Unknown
12	83.33	yes		no	0.00	Unknown	no	no	0.00	Unknown
11	90.91	yes		no	0.00	Unknown	no	no	0.00	Unknown
									1	•

Figure 15: You can filter for zebu specific variants in the overview table.

6. Click on the button "Filter", select all remaining rows and click on "Create Track from Selection". Save the new track in the Navigation Area.

The output from this analysis is a new variant track that holds only the variants (less than 3000) that are specific for the zebu species Bos indicus called "(Reads, Locally Realigned, Variants) (MutationTest overview)\_selection". You can also see that you have different types of variation, some of which are heterozygous and others homozygous. You can export tables from the Workbench in Excel formats or text formats. Click on the table so that the table you would like to export is the selected view, then use the **Export** ( $\square$ ) button in the toolbar to start the export. For small tables, you can also just **Copy** ( $\square$ ) the contents of the table and paste into a spreadsheet for further processing.

We will now look only at the variants that are present in the coding part of the genome. We can do this with the tool **Filter Based on Overlap**.

1. Go to:

#### Toolbox | Track Tools ((a) | Annotate and Filter (a) | Filter Based on Overlap ()

- 2. In the first wizard step, select the zebu specific variant track that you just saved as input. Click **Next**.
- 3. Select the **Bos taurus (Hereford)\_CDS** track for overlap comparison and **Keep annotations that overlap**. Click **Next**.
- 4. Choose to save the results and click on the button labeled Finish.

The result is a variant track called "(Reads, Locally Realigned, Variants) (MutationTest overview)\_selection (OF)" and containing a number in the range of 600 zebu-specific variants that occurs in the CDS regions.

Now we would now like to look only at variants that cause amino acid changes. We can do this with the tool **Amino Acid Changes**.



1. To run the analysis, go to the toolbox:

Toolbox | Resequencing Analysis () | Functional Consequences () | Amino Acid Changes ()

- 2. In the first wizard step, select the most recent zebu specific variant track (the one filtered based on overlap with CDS) as input. Click **Next**.
- 3. Select the CDS, mRNA, and sequence tracks as shown in figure 16 and choose to detect only non-synonymous changes by ticking the box "Filter synonymous". Leave also the option "Filter CDS regions with no variants" checked. Click **Next**.

Gx Amino Acid Changes		x
1. Select variant track	Set parameters	
2. Set parameters	Select CDS track CDS track CDS track CDS taurus (Hereford)_CDS	a l
	Select transcript (mRNA) track	
	mRNA track 🚓 Bos taurus (Hereford)_mRNA	ିଇ
	Select sequence track	
Col man	Sequence track 🎇 Bos taurus (Hereford) sequence	ିଇ
(22)	Filtering and annotation	
	✓ Filter synonymous variants	
The Contraction of the Contracti	Filter CDS regions with no variants Genetic code Standard	
Louise Harris	Genetic code 1 Standard V	
LEALDALATIN		
? 4	← Previous → Next ✓ Finish X Ca	incel

Figure 16: Select CDS, mRNA, and sequence tracks and check Filter synonymous and Filter CDS regions with no variants.

4. Choose to save the results and click on the button labeled **Finish**.

The result is a variant track called "(Reads, Locally Realigned, Variants) (MutationTest overview)\_selection (OF, AAC)" with over 200 non-synonymous zebu-specific variants found in the CDS regions that cause amino acid changes.

#### **Overrepresentation analysis**

To gain more insight into the underlying biology we are now going to perform a gene ontology overrepresentation analysis with the tool **GO Enrichment Analysis**. The Gene Ontology site http://www.geneontology.org/GO.downloads.annotations.shtml contains the bovine gene ontology annotation.

- Go to http://cvsweb.geneontology.org/cgi-bin/cvsweb.cgi/go/gene-association gene\_association.goa\_cow.gz?rev=HEAD and download the bovine gene ontology annotation to your computer (1.6 MB).
- 2. Import the file into the Workbench with the import function found in the toolbar: Import | Standard Import .



- 3. Select **Gene Ontology Annotation File** as **Format** in the import wizard (or alternatively use **Automatic Import**, which will give the same result). The file does not have to be unzipped before import.
- 4. To analyze if any biological pathways are overrepresented for the detected zebu-unique variants, go to the toolbox and run the tool:

Toolbox | Resequencing Analysis () | Functional Consequences () | GO Enrichment Analysis ()

- 5. In the first wizard step, select the variant track that was created with the Amino Acid Changes tool Bos indicus (Reads) locally realigned (Variants, CTRL, OF, AAC) as input. Click Next.
- 6. Select the parameter track **Bos taurus (Hereford)\_Gene** and the GO annotation table, uncheck **Exclude computationally inferred GO terms**, and check **Allow gene name synonyms** as shown in figure 17. Click **Next**.

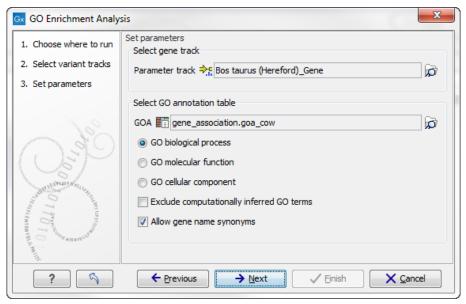


Figure 17: Select parameter track and GO annotation table as specified.

7. Choose to save the results and click on the button labeled **Finish**.

You get two different types of output: a **GO-enrichment table** and a variant track that is essentially the same variant table as the one generated in the **Amino Acid Changes** analysis but with five extra columns (see the red box in the side panel in figure 18).

Open the GO-enrichment table by clicking on the file name in the **Navigation Area**. We can sort the rows by the calculated p-value (from a hypergeometric distribution test comparing frequency of occurrence in a sample vs in all genes) by clicking once on the column header "P-values". It is clear that in the top rows the pathways related to the immune response dominate (figure 19). This matches previous findings showing that zebu responds differently to some infections and also has a better innate immune response compared to the taurine and the bison species [Freeman et al., 2008].



-

I⊧ Ta	able Settings
	Amino acid change in longest transcript
	Coding region change in longest trans
	Other variants within codon
	Non-synonymous
<u>ا</u>	Overlapping gene name
	Overlapping gene names w/GO-term
	GO-terms
	GO-terms descriptions
	P-values
	Select All
	Deselect All

Figure 18: Five extra columns have been added compared to the Amino Acid Changes table.

Rows: 7,994	GO enrichment analysis			Filter
GO term	Description	Occurrences in all genes	Occurrences in sample	P-values /
032757	positive regulation of interleukin-8 production	16	5	1.11E
002755	MyD88-dependent toll-like receptor signaling pathway	9	4	1.80E
034123	positive regulation of toll-like receptor signaling pathway	4	3	3.10E
032755	positive regulation of interleukin-6 production	34	6	4.99E
006955	immune response	137	11	9.16E
042346	positive regulation of NF-kappaB import into nucleus	14	4	1.32E
007252	I-kappaB phosphorylation	7	3	2.59E
050707	regulation of cytokine secretion	7	3	2.59E
042496	detection of diacyl bacterial lipopeptide	2	2	3.97E
071726	cellular response to diacyl bacterial lipopeptide	2	2	3.97E
002504	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	10	3	8.50E
032722	positive regulation of chemokine production	11	3	1.15E
032611	interleukin-1 beta production	3	2	1.18E
045356	positive regulation of interferon-alpha biosynthetic process	3	2	1.18E
071224	cellular response to peptidoglycan	3	2	1.18E
050729	positive regulation of inflammatory response	25	4	1.41E
044130	negative regulation of growth of symbiont in host	12	3	1.51E
045087	innate immune response	89	7	1.98E
001867	complement activation, lectin pathway	4	2	2.32E
042773	ATP synthesis coupled electron transport	4	2	2.32E
019221	cytokine-mediated signaling pathway	32	4	3.59E
002024	diet induced thermogenesis	5	2	3.82E
002830	positive regulation of type 2 immune response	5	2	3.82E
042116	macrophage activation	5	2	3.82E
008203	cholesterol metabolic process	35	4	4.99E
002925	positive regulation of humoral immune response mediated by circulating immunoglobulin	6	2	5.65E
045359	positive regulation of interferon-beta biosynthetic process	6	2	5.65E

Figure 19: Sort on p-value by clicking once on the column header.

#### Automating the analysis as workflows

Some of the steps in this analysis can be combined into workflows. This is shown in the two examples in (figure 20 and figure 21). These workflows, combined with batch mode of operation, greatly simplify the analysis, and can be used to repeat it, for instance, to look for similar variant enrichment in the other two species of bovines. For further information about how to create and run workflows, please see the tutorial "An Introduction to Workflows" that can be found here: http://resources.giagenbioinformatics.com/tutorials/Workflow-intro.pdf.



		🛛 🕪 Input rea	ids		
		<u> </u>			
	Reads	Reference	5 N	Asking track	
	📷 Map Rea	ds to Reference			
	Reads Track	Read Mapping	Mapping Repo	ort Un-mappe	ed Reads
				<pre></pre>	
Read Mapping or	r Reads Track	Guidance-varian	t track		
Local Realig	nment			🕼 Mappin	g Report
Reads Track	Read Mapping	Realigned Re			
Reads Track	Read Mapping	Realigned Re	gions		
			-		
		📃 🤃 Re	ads Track		
ead Mapping or	Reads Track	Restrict calling to	target regions		
<b>A</b>					
A Basic Variant	t Detection				
ariant Track		Variant Table			
	×.				
	🖕 🕼 Variant Tra	ck · · · · ·			

Figure 20: Mapping of the reads and variant calling workflow.

Input Track Filter E Annotation	Based or	Variant Track	Expr	Overlap trac		al Comp	arison Tra	
Filter E	Based on Track	Variant Track	Expr			al Comp	arison Tra	_
Annotation	Track	Variant Track	Expr	ession Track	Statistic	al Comp	arison Tra	_
	K		Expr	ession Track	Statistic	al Comp	arison Tra	ick
	Variant T							
	Variant T							
		rack CDS trac	sk S	equence track	mRNA	A track		
• • • • •	Amino Acid Chan		jes 🖹			Ē		
Amino Aci		cid Track Annotated Varian		nt Track				
· · · · · · · ·								
		Variant Track	Para	ameter track	GOA			
	GO Enrichm		ment A	Inalysis	Ê			
		Annotated Trac	k	Output Tab	le			
· · · · · · · · ·			· · ·					
	Ann	otated Track		📋 🥠 Out	put Table	2		

Figure 21: Variant filtering and annotation workflow.



## **Bibliography**

- [Cosart et al., 2011] Cosart, T., Beja-Pereira, A., Chen, S., Ng, S. B., Shendure, J., and Luikart, G. (2011). Exome-wide dna capture and next generation sequencing in domestic and wild species. *BMC Genomics*, 12:347.
- [Freeman et al., 2008] Freeman, A. R., Lynn, D. J., Murray, C., and Bradley, D. G. (2008). Detecting the effects of selection at the population level in six bovine immune genes. *BMC Genet*, 9:62.